NEURIPS 2026

SkillForge: Co-Evolving Skills and Agents via Dynamic Skill Lifecycles

Yuyao Ge[1,2,3] · Yiwei Wang[4] · Yuchen He[1,2,3] · Baolong Bi[1,2,3] · Lingrui Mei[1,2,3] · Jiayu Yao[1,2,3] · Lizhe Chen[5] · Shenghua Liu[1,2,3,†]

[1]State Key Laboratory of AI Safety [2]Institute of Computing Technology, CAS [3]University of Chinese Academy of Sciences
[4]University of California, Merced [5]Tsinghua University [†]Corresponding author



PAPER — arXiv 2610.09832 · CODE — GitHub · PAGE — Project page

TL;DR An append-only skill library goes stale as the policy improves. **SkillForge** gives every skill a fitness score and a lifecycle — `trial` → `active` → `stable` → `retired` — and lets a teacher model rewrite borderline skills, so **the skill library and the agent co-evolve under RL**.

01 · PROBLEM

A library that only grows gets dragged down by stale skills

A **skill** is a kind of agent memory: an instruction paired with an applicability condition, retrieved into the agent's context when it applies.

SKILL **Pick Before You Wander**

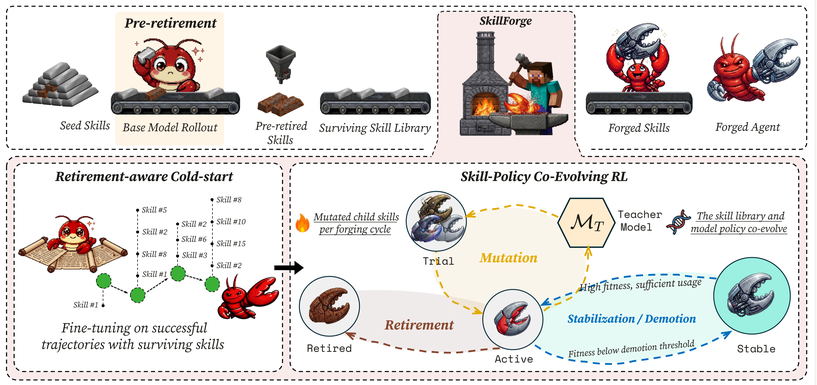
Instruction	*If the target object is visible or discovered, take it immediately; never leave it behind to explore other places.*
Applies to	retrieved by task type (ALFWorld)

Most skill-augmented RL methods, such as SkillRL, only **append** to the library. As the policy improves, outdated and even harmful skills keep piling up and keep being retrieved into context.

Delayed obsolescence. A skill that once helped can later hold the agent back. The skills we retired had peaked at an average fitness of 0.67, close to the stable threshold of 0.7, and had lost 0.28 by the time they were retired; 57.2% of them had once been `stable`.

02 · METHOD

Filter out weak skills, then co-evolve skills and policy



1. **Pre-retirement.** Roll out the base model with the seed skills and drop every skill whose proto-fitness $\hat{f}_i = C_i^{(0)}/U_i^{(0)}$ is below 0.3.
2. **Retirement-aware cold start.** Fine-tune only on successful trajectories that use surviving skills.
3. **Skill–policy co-evolving RL.** Train with GRPO; every 10 steps a *forging cycle* retires, stabilizes and mutates skills by fitness.

Borderline skills are rewritten by a teacher

`active` skills with $0.4 \le f_i \le 0.7$ are sampled with $P(s_i) \propto 1 - f_i$. A teacher model (Kimi-K2.5) reads their failure trajectories and writes a child skill, which enters the library as `trial`:

$$s_{\text{child}} = \mathcal{M}_T\big(\mathcal{P}_\mu(s_{\text{parent}},\ f_{\text{parent}},\ \mathcal{T}_{\text{fail}})\big)$$

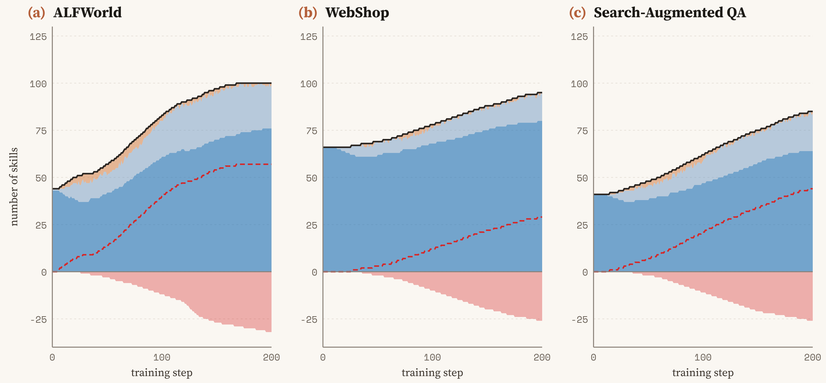
02 · METHOD · SKILL LIFECYCLE

Every skill moves through four states by its fitness

Each episode retrieves the top-6 skills matching its task type. A skill's fitness is its success rate $f_i = C_i/U_i$ over the U_i episodes that retrieved it.

- `trial` → `active` after $U_i \ge 10$ uses
- `active` → `stable` $f_i \ge 0.7$ and $U_i \ge 30$
- `stable` → `active` $f_i < 0.5$
- `active` → `retired` $f_i < 0.4$ and $U_i \ge 20$

The library levels off while skills keep turning over



- **The library stays compact.** On ALFWorld it grows to 100 skills (132 created, 32 retired); once the cap binds near step 170, mutation and retirement balance out.
- **Most skills end up stable:** 76 of 100 on ALFWorld, 80 of 95 on WebShop and 64 of 85 on Search QA.
- **Same pattern on all three benchmarks:** retirements keep accumulating to the last step while mutated children take their place.

Every part matters, and more skills is not better

Ablation: library size (#) and success / accuracy (%).

Configuration	ALFWorld		WebShop		Search QA	
	#	Succ.	#	Succ.	#	Acc.
SkillForge	100	**92.4**	95	**78.4**	85	**48.7**
w/o forging	44	89.8	66	72.6	41	46.3
w/o pre-retirement	100	90.7	95	74.9	85	47.5
w/o retirement	**132**	91.4	**121**	75.9	**111**	47.9
w/o mutation	44	91.8	66	76.8	41	48.2

- **Removing the whole lifecycle hurts most** on every benchmark, by up to 7.4% relative on WebShop.
- **Pre-retirement is the most important single part:** relative drops of 1.8%, 4.5% and 2.5% without it.
- **More skills is not better:** without retirement the library grows to 132 skills on ALFWorld, yet does worse than SkillForge with only 100.

TAKEAWAY

SkillForge turns skill-library management into a fitness-driven forging process, so that **skills and the model co-evolve**.

03 · EXPERIMENTS · MAIN RESULTS

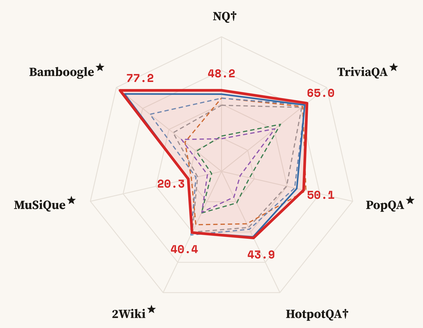
Beats the strongest baseline on all three benchmarks

ALFWorld and WebShop (%). Score: partial-credit reward. **Bold**: best; underline: second best.

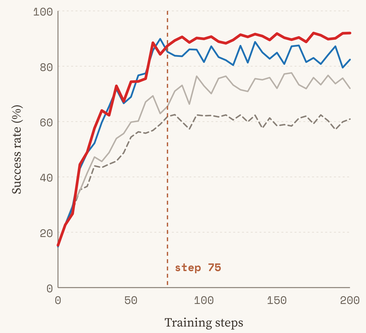
Method	ALFWorld							WebShop	
	Pick	Look	Clean	Heat	Cool	Pick2	All	Score	Succ.
Closed-source LLMs									
GPT-4o	75.3	60.8	31.2	56.7	21.6	49.8	48.0	31.8	23.7
Gemini-2.5-Pro	92.8	63.3	62.1	69.0	26.6	58.7	60.3	42.5	35.9
Base Model									
Vanilla	33.4	21.6	19.3	6.9	2.8	3.2	14.8	26.4	7.8
Context / Memory Methods									
ReAct	48.5	35.4	34.3	13.2	18.2	17.6	31.2	46.2	19.5
Reflexion	62.0	41.6	44.9	30.9	36.3	23.8	42.7	58.1	28.8
Mem0	54.0	55.0	26.9	36.4	20.8	7.7	33.6	23.9	2.0
ExpeL	21.0	67.0	55.0	52.0	71.0	6.0	46.3	30.9	11.2
MemP	54.3	38.5	48.1	56.2	32.0	16.7	41.4	25.3	6.4
SimpleMem	64.5	33.3	20.0	12.5	33.3	3.8	29.7	33.2	8.6
RL-based Methods									
RLOO	87.6	78.2	87.3	81.3	71.9	48.9	75.5	80.3	65.7
GRPO	90.8	66.1	89.3	74.7	72.5	64.7	77.6	79.3	66.1
Memory-Augmented RL									
MemRL	62.8	38.5	22.2	12.5	8.0	0.0	21.4	29.5	9.2
EvolveR	64.9	33.3	46.4	13.3	33.3	33.3	43.8	42.5	17.6
Mem0+GRPO	78.1	54.8	56.1	31.0	65.0	26.9	54.7	58.1	37.5
SimpleMem+GRPO	89.5	36.3	60.0	50.0	64.9	26.3	62.5	67.8	46.9
SkillRL	97.9	71.4	**90.0**	90.0	95.5	87.5	89.9	85.2	72.7
SkillForge	**98.2**	**86.6**	89.8	**93.7**	**95.6**	**88.4**	**92.4**	**87.8**	**78.4**

- **A 7B model beats frontier LLMs:** on Qwen2.5-7B-Instruct, SkillForge far exceeds GPT-4o and Gemini-2.5-Pro.
- **Ahead of SkillRL, the strongest baseline:** +2.8% relative on ALFWorld and +7.8% on WebShop; best in 8 of the 9 columns.
- **Largest subtask gain on Look** (+21.3% relative over SkillRL); Clean stays within 0.2 of SkillRL.

(a) Search-Augmented QA (%)



(b) ALFWorld validation success rate



SkillForge (ours) · SkillRL · EvolveR · ZeroSearch · Search-R1 · Search-o1 · R1-Instruct · Vanilla GRPO · GRPO+Memory · † in-domain · ★ out-of-domain

- **Highest overall accuracy:** 48.7 vs. 46.8 for SkillRL, over 51,713 questions.
- **Best or second best on all seven datasets**, both in-domain and out-of-domain.
- **Largest gain on Bamboogle:** 77.2 vs. 73.8.
- **Both rise together** until about step 65.
- **Forging pays off after step 75:** SkillForge reaches ~90%; SkillRL oscillates at ~80–88%.
- Vanilla GRPO and GRPO+Memory level off far lower.